

IFT 187 – THÈME 2

NORMALISATION

Nadia Tahiri, Ph. D.
Professeure adjointe
Université de Sherbrooke

Nadia.Tahiri@USherbrooke.ca



DÉPENDANCE FONCTIONNELLE

Source: Marc Frappier (UdeS), IFT187



DÉPENDANCE FONCTIONNELLE

- les définitions de 2FN, 3FN, BCNF reposent sur la notion de dépendance fonctionnelle
- une **dépendance fonctionnelle** est une **fonction** entre des listes d'attributs
- on dénote une dépendance fonctionnelle comme suit :

$$(A_1, \dots, A_n) \rightarrow A_{n+1}$$

on dit que A_{n+1} dépend de A_1, \dots, A_n

QUE REPRÉSENTE UNE DÉPENDANCE FONCTIONNELLE?

- C'est une fonction, donc elle associe à une liste de valeurs des attributs A_1, \dots, A_n une et une seule valeur dans A_{n+1}
- exemple
 - dans une université, étant donné le matricule d'un étudiant, on peut donner son nom
 - il existe donc une dépendance fonctionnelle entre matricule et nom

matricule \rightarrow nom

- **l'inverse n'est pas vrai** : étant donné un nom, on ne peut déterminer le matricule d'un étudiant, car il peut y avoir plusieurs matricules, puisque plusieurs étudiants peuvent avoir le même nom

QUE REPRÉSENTE UNE DÉPENDANCE FONCTIONNELLE?

- **Attention!**

la dépendance

matricule \rightarrow *nom*

- **ne signifie pas** que le nom associé à un matricule ne change jamais; le nom peut changer, mais, en tout temps, on peut déterminer le nom d'un étudiant à partir de son matricule
- cela **ne signifie pas non plus** que si on a deux matricules différents, alors leurs noms associés doivent être différents
- cela signifie que deux étudiants ne peuvent avoir le même matricule

DÉPENDANCE FONCTIONNELLE MINIMALE

- si

$$(A_1, \dots, A_n) \rightarrow B$$

alors on a aussi

$$(A_1, \dots, A_n, A_{n+1}) \rightarrow B$$

- pour les fins de normalisation, on considère seulement les dépendances qui sont minimales selon la liste de gauche

DÉPENDANCE FONCTIONNELLE ET CLÉ CANDIDATE

- s'il existe une dépendance fonctionnelle minimale entre (A_1, \dots, A_n) et tous les autres attributs de la relation, alors on peut conclure que (A_1, \dots, A_n) est une clé candidate
- une dépendance fonctionnelle sera donc traduite en une contrainte `primary key` ou `unique`

QUELQUES LOIS SUR LES DÉPENDANCES FONCTIONNELLES



Soit W , X , Y et Z des ensembles d'attributs

Par soucis de concision, on dénote par XY

l'union $X \cup Y$ de deux ensembles d'attributs X et Y

l'inclusion $X \subseteq Y$ de deux ensembles d'attributs X et Y

1. Si $X' \subseteq X$, alors $X \rightarrow X'$ (DF triviale)
2. Si $X \rightarrow Y$, alors $X \cup Z \rightarrow Y$ (augmentation)
3. Si $X \rightarrow Y$ et $Y \rightarrow Z$, alors $X \rightarrow Z$ (transitivité)
4. Si $X \rightarrow Y \cup Z$, alors $X \rightarrow Y$ et $X \rightarrow Z$ (décomposition)
5. Si $X \rightarrow Y$ et $X \rightarrow Z$, alors $X \rightarrow Y \cup Z$ (union)
6. Si $X \rightarrow Y$ et $W \cup Y \rightarrow Z$, alors $X \cup W \rightarrow Z$ (pseudo-transitivité)

COMMENT DÉTERMINER LES DÉPENDANCES FONCTIONNELLES?



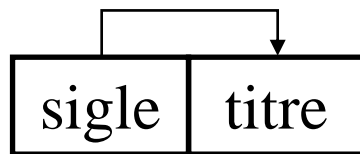
- les dépendances fonctionnelles sont des contraintes du domaine d'application
- on les détermine à partir de notre connaissance des faits (règles, conditions, etc) du domaine d'application
- on peut déterminer s'il y a une dépendance fonctionnelle

$(A_1, \dots, A_n) \rightarrow A_{n+1}$ en répondant à la question suivante:

- étant donné une liste de valeurs pour A_1, \dots, A_n , peut-on toujours associer une et une seule valeur pour A_{n+1} ?

REPRÉSENTATION GRAPHIQUE

- $\text{sigle} \rightarrow \text{titre}$



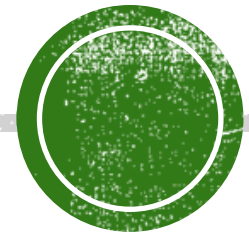
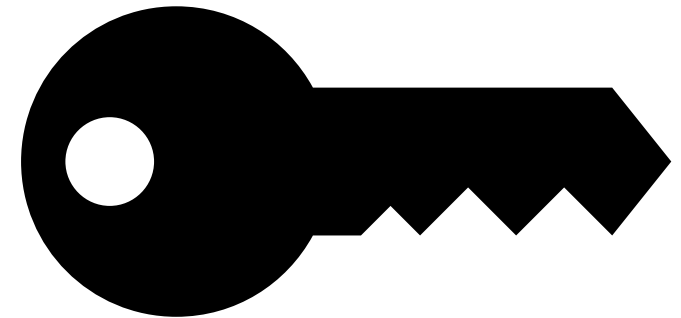
- $(\text{sigle}, \text{session}, \text{groupe}) \rightarrow \text{matricule}$



EXERCICE

- identifiez les dépendances fonctionnelles entre les attributs suivants
 - sigle, titre, matricule, nom, session, groupe, note, salaire, coteR

DEUXIÈME FORME NORMALE (FN2)



DEUXIÈME FORME NORMALE (FN2)

- Une table respecte la 2ième forme normale si:
 - Elle respecte la **FN1**
 - Tout attribut ne faisant pas partie de la clé est complètement dépendant de la clé primaire (ne dépend pas d'une partie de la clé)
- Cette forme normale ne s'applique que dans le cas de clés composées (multi-colonnes).

DEUXIÈME FORME NORMALE (FN2)

- **Exemple non FN2**

Souligné : Clé

Table: Employe_Projet

<u>id_employe</u>	<u>id_projet</u>	role	nom_employe	nom_projet
10	45	Programmeur	Lapierre, A.	Caisse de dépôt
12	37	Designer FrontEnd	Desjardins, R.	Therac
10	33	Architecte	Lapierre, A.	Geothermia
18	37	Concepteur	Labonté, T.	Therac

DEUXIÈME FORME NORMALE (FN2)

- Table normalisée en FN2

Table: Employe

<u>id_employe</u>	nom
10	Lapierre, A.
12	Desjardins, R.
18	Labonté, T.

Table: Projet

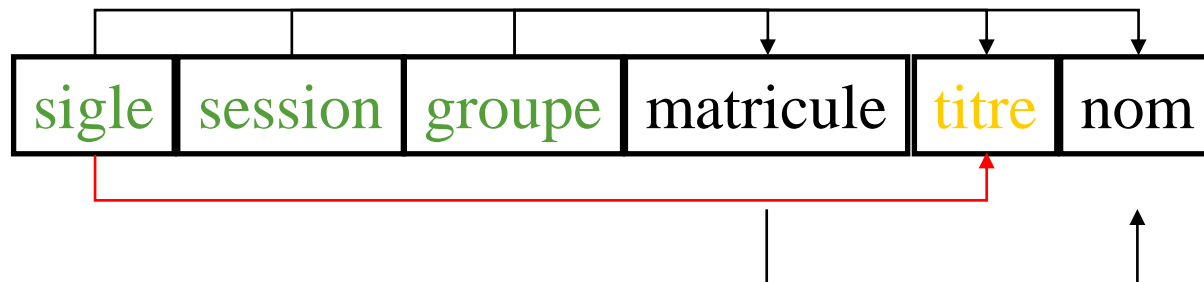
<u>id_projet</u>	nom
45	Caisse de dépôt
33	Geothermia
37	Therac

Table: Employe_Projet

<u>id_employe</u>	<u>id_projet</u>	role
10	45	Programmeur
12	37	Designer FrontEnd
10	33	Architecte
18	37	Concepteur

CONTRE-EXEMPLE DE FN2

Une entité E est en deuxième forme normale ssi tous les attributs *non premiers* de E sont en *dépendance fonctionnelle complète* de chaque *clé candidate* de E

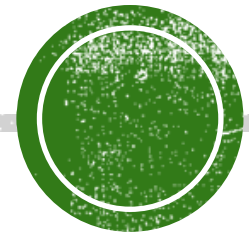
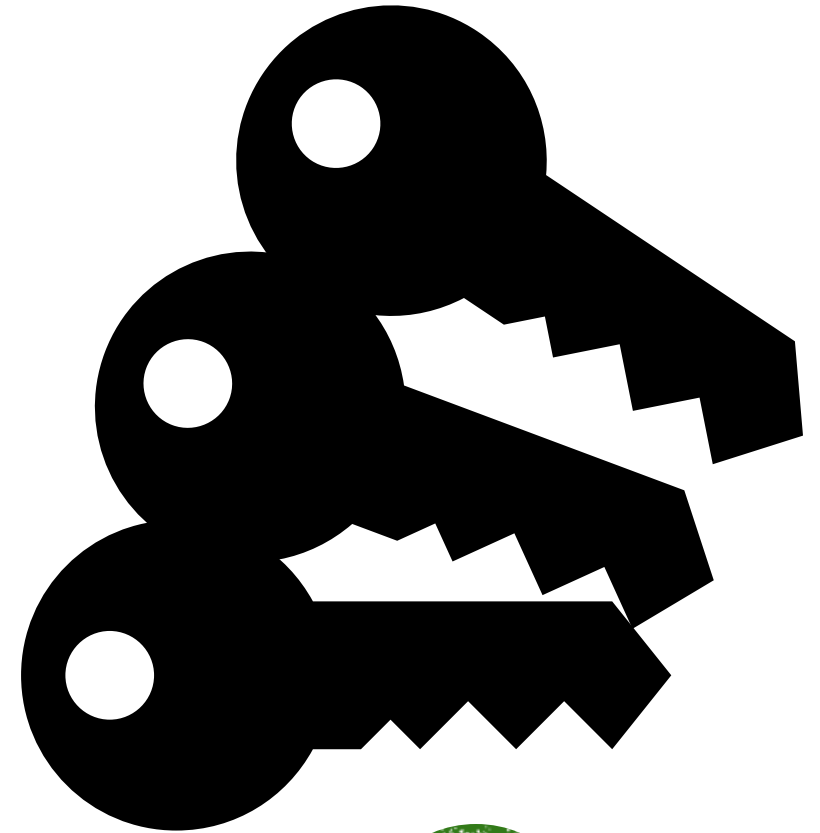


titre ne dépend pas de toute la clé; il dépend seulement de *sigle*

POURQUOI NORMALISER EN FN2 ?

- ✓ parce que cela élimine la redondance des données
- ✓ cela assure une meilleure intégrité des données tout en simplifiant les mise à jour
- ✓ on ne perd aucune information; on peut recréer la relation originale avec une jointure des deux relations normalisées

TROISIÈME FORME NORMALE (FN3)



TROISIÈME FORME NORMALE (FN3)

- Une table est en troisième forme normale si:
 - Elle respecte la FN2
 - Toute colonne non-clé dépend non-transitivement de la clé primaire.
(i.e. ne dépend pas d'un ou plusieurs attributs n'appartenant pas à la clé).
- Ou...

Une relation E est en troisième forme normale ssi pour toute dépendance fonctionnelle $X \rightarrow A$ de E, une des conditions suivantes est satisfaite:

- X est une **super clé**
- A est un attribut **premier**

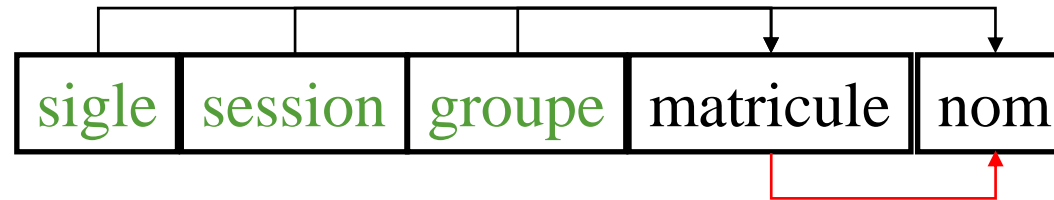
QUELQUES DÉFINITIONS

- **Super-clé**
 - Ensemble d'attributs X qui sont uniques à chaque tuple.
- **Clé (ou clé candidate)**
 - Super-clé minimale.
 - Super-clé X telle que si on enlève n'importe quel attribut, ce n'est plus une super-clé.
- **Attribut premier** (parfois appelé primaire)
 - Un attribut qui fait partie d'au moins une clé
 - Si X est un ensemble d'attributs, il est premier si au moins un de ses attributs est premier.
 - Attribut non-premier = un attribut qui n'est pas premier

CONTRE-EXEMPLE DE FN3

Une relation E est en troisième forme normale ssi pour toute dépendance fonctionnelle $X \rightarrow A$ de E, une des conditions suivantes est satisfaite :

- X est une **super clé**
- A est un attribut **premier**

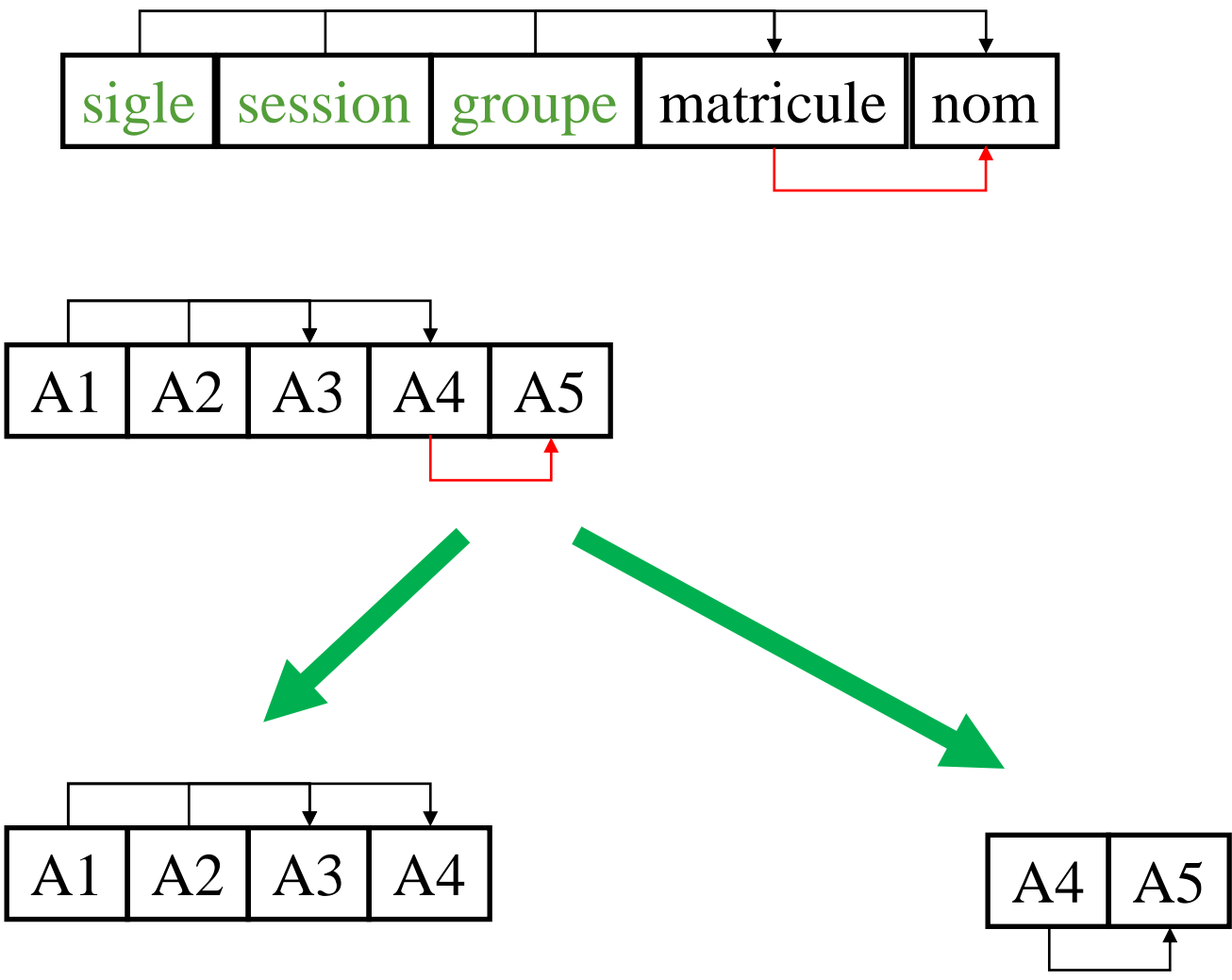


cette relation n'est pas FN3 car :

- *matricule* n'est pas une **super clé**
- *nom* n'est pas **premier**

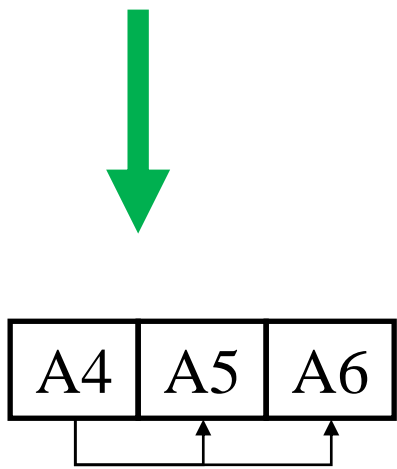
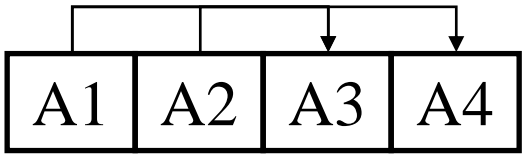
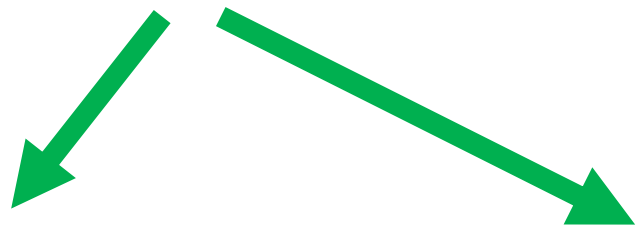
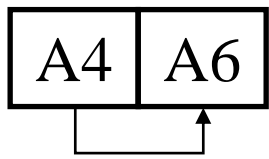
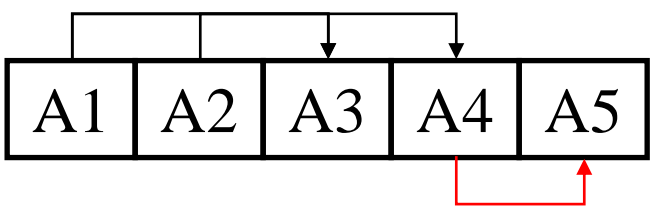
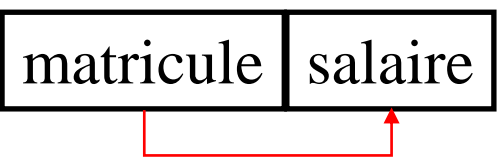
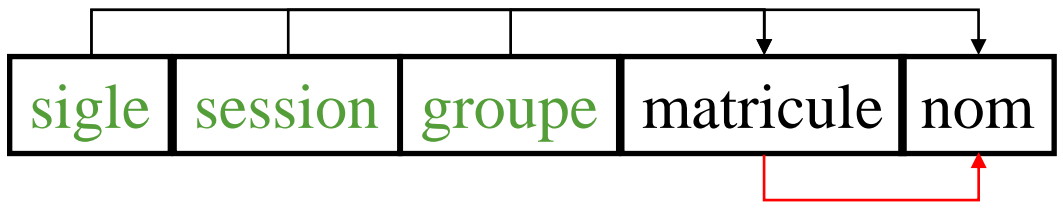
NORMALISATION EN FN3

Scénario 1



NORMALISATION EN FN3

Scénario 2

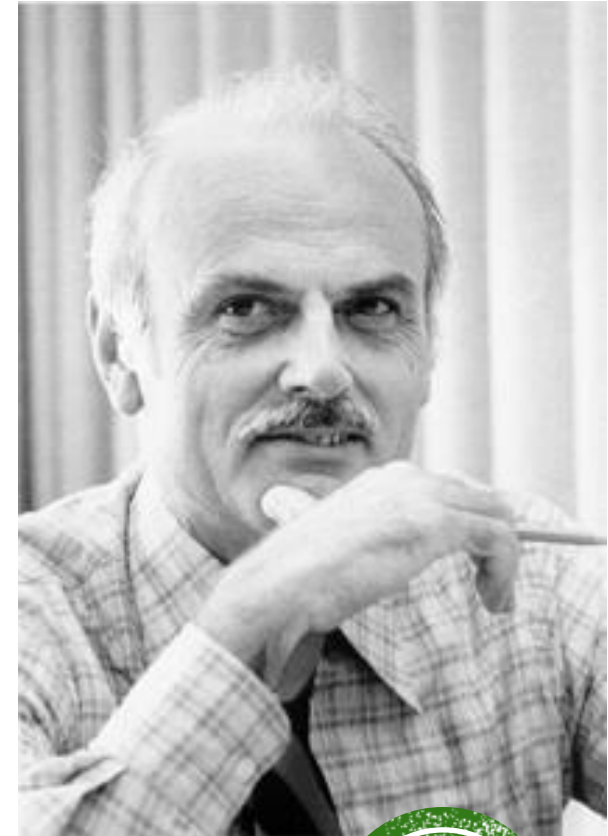


POURQUOI NORMALISER EN 3FN

- comme pour la 2FN
 - parce que cela élimine la redondance des données
 - cela assure une meilleure intégrité des données tout en simplifiant les mise à jour
- on ne perd aucune information; on peut recréer la relation originale avec une jointure des deux relations normalisées

FORME NORMALE BOYCE-CODD (FNBC)

Edgar F. Codd (1923-2003)



FORME NORMALE BOYCE-CODD (FNBC)

- Une table est en forme normale Boyce-Codd (FNBC) si:
 - Elle respecte la FN3
 - Tous les attributs non-clé ne sont pas source de dépendance vers une partie de la clé.

Ou

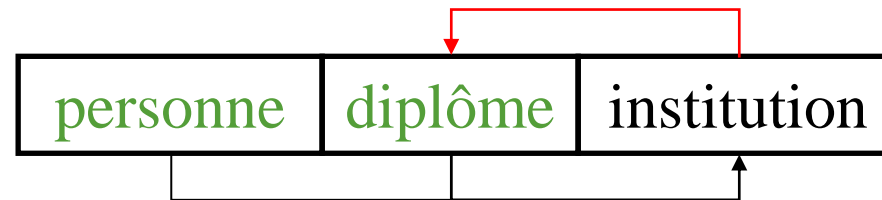
- Une relation E est en forme normale de Boyce-Codd ssi pour toute dépendance fonctionnelle

$X \rightarrow A$ de E, la condition suivante est satisfaite :

- X est une **super clé**

CONTRE-EXEMPLE DE BCNF

- supposons qu'une institution d'enseignement décerne un seul diplôme (SEC, DEC, ou BAC) et qu'une personne obtient un diplôme d'une et une seule institution; on a les DF suivantes
 - $(personne, diplôme) \rightarrow institution$
 - $institution \rightarrow diplôme$

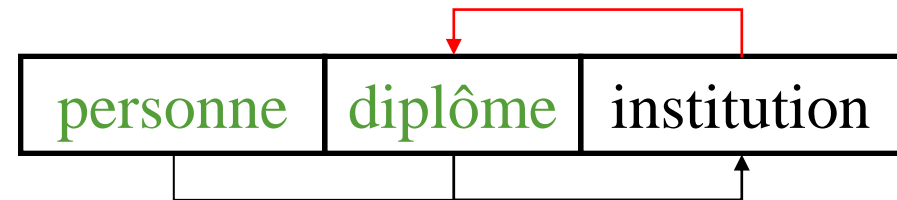


CONTRE-EXEMPLE DE BCNF

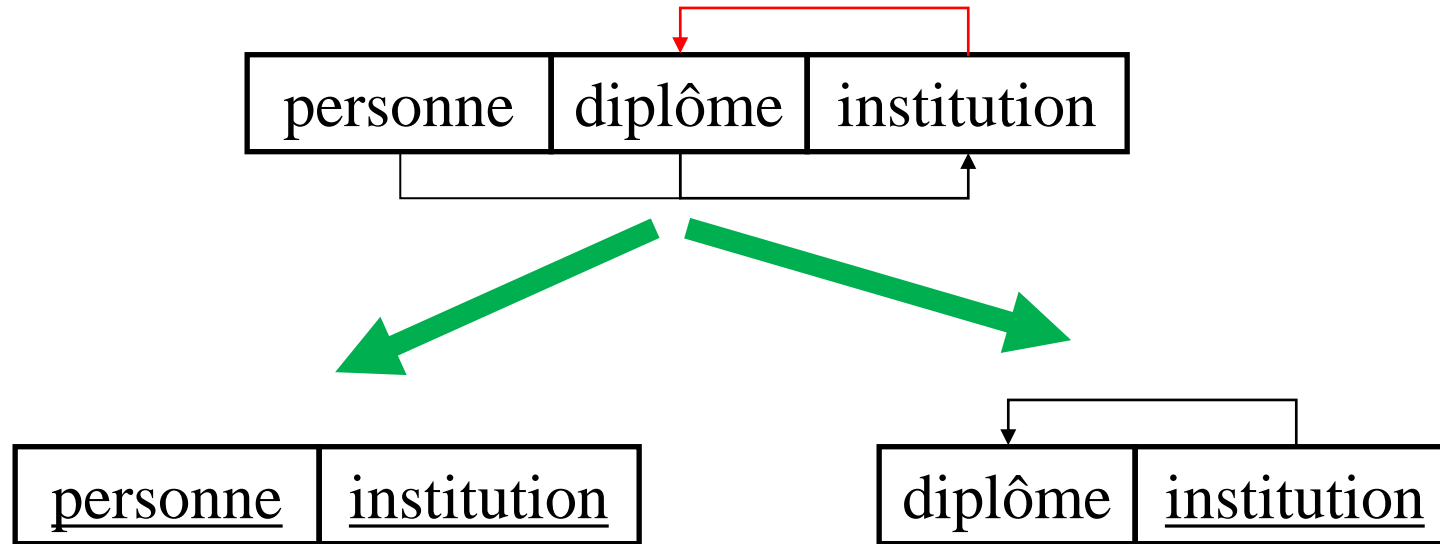
Une entité E est en forme normale de Boyce-Codd ssi pour toute dépendance fonctionnelle $X \rightarrow A$ de E, la condition suivante est satisfaite:

–X est une **super clé**

Cette entité n'est pas en BCNF, car il y a la DF *institution* \rightarrow *diplôme*, et *institution* n'est pas une **super clé**



NORMALISATION EN BCNF



Note :

- on ne perd pas d'information,
- on diminue la redondance
- on perd une contrainte d'intégrité
(*personne, diplôme*) \rightarrow *institution*

QUATRIÈME FORME NORMALE (FN4)

Source: Robert Godin, 2012



QUATRIÈME FORME NORMALE ET DÉPENDANCES MULTIVALUÉES

- *Film* est en **FNBC** mais pas en **4FN**
- Redondance due à dépendance multivaluée
 - *titre* → *nomProducteur*
 - *titre* → *nomActeur*

Table <i>Film</i>		
titre	nomProducteur	nomActeur
La vie est belle	Elda Ferri	Roberto Benigni
La vie est belle	Elda Ferri	Nicoletta Braschi
La vie est belle	Elda Ferri	Giorgio Cantarini
La vie est belle	Gianluigi Braschi	Roberto Benigni
La vie est belle	Gianluigi Braschi	Nicoletta Braschi
La vie est belle	Gianluigi Braschi	Giorgio Cantarini
Patch Adams	Barry Kemp	Robin Williams
Patch Adams	Barry Kemp	Monica Potter
Patch Adams	Michael Farrell	Robin Williams
Patch Adams	Michael Farrell	Monica Potter
Patch Adams	Marvin Minoff	Robin Williams
Patch Adams	Marvin Minoff	Monica Potter

DÉPENDANCE MULTIVALUÉE (MULTIVALUED DEPENDENCY)

- $A_1, A_2, \dots, A_n \twoheadrightarrow B_1, B_2, \dots, B_m$
 - à chacune de valeurs de A_1, A_2, \dots, A_n est associé un ensemble de valeurs de B_1, B_2, \dots, B_m sans relation directe avec les autres colonnes C_1, C_2, \dots, C_p de la table

DÉFINITION FORMELLE

- $(a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m, c_1, c_2, \dots, c_p)$ et $(a_1, a_2, \dots, a_n, b'_1, b'_2, \dots, b'_m, c'_1, c'_2, \dots, c'_p) \in T$
 $\Rightarrow (a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m, c'_1, c'_2, \dots, c'_p) \in T$
- ('La vie est belle', 'EldaFerri', 'Robert Benigni') et ('La vie est belle', 'Gianluigi Braschi', 'Nicoletta Braschi') $\in Film \Rightarrow$
('La vie est belle', 'EldaFerri', 'Nicoletta Braschi')
 $\in Film$

DÉPENDANCES MULTIVALUÉES

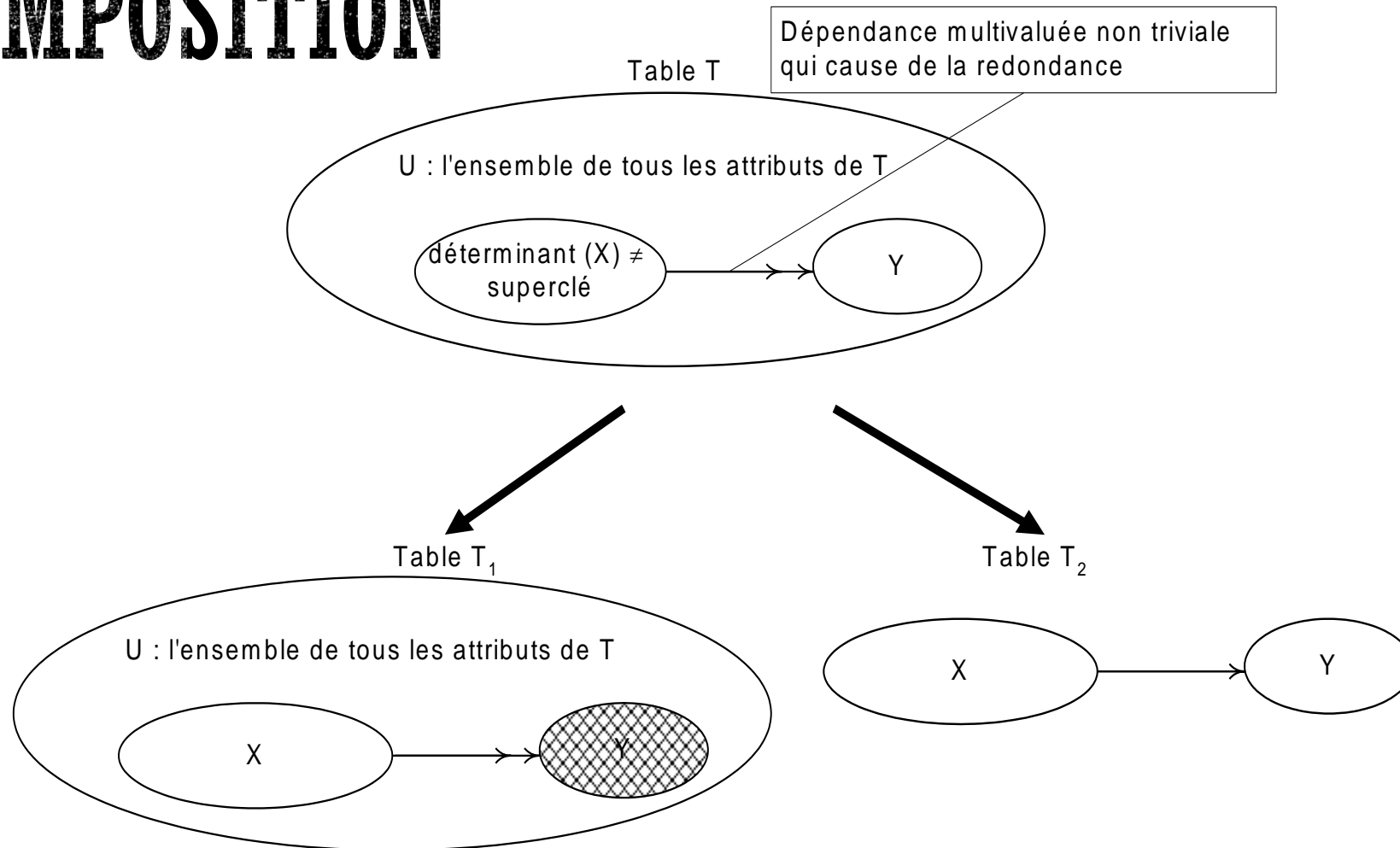
- **Proposition** (*règle du complément*)

- $A_1, A_2, \dots, A_n \twoheadrightarrow B_1, B_2, \dots, B_m \Rightarrow A_1, A_2, \dots, A_n \twoheadrightarrow C_1, C_2, \dots, C_p$
 - C_1, C_2, \dots, C_p correspond aux autres colonnes de T

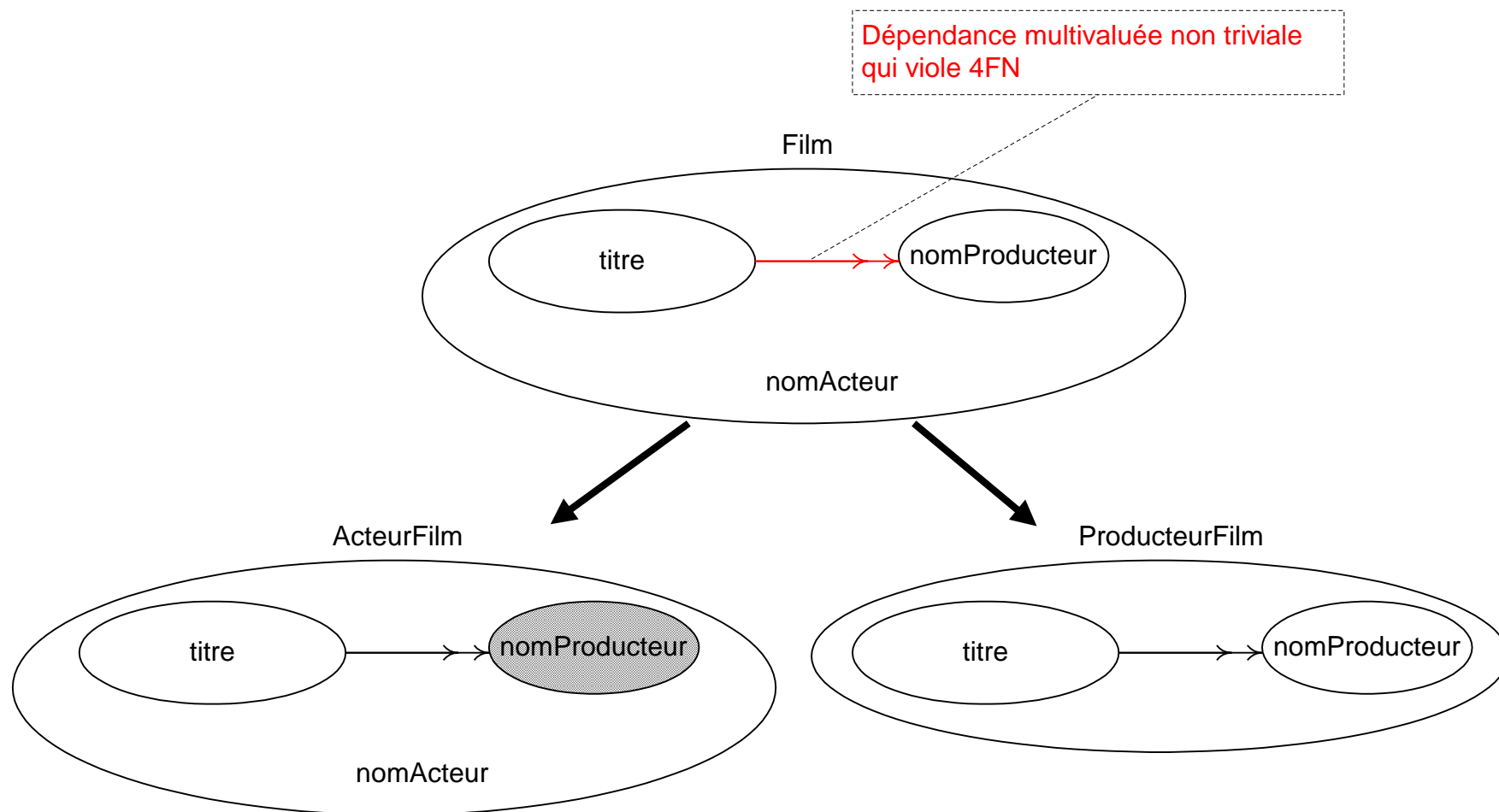
- **Proposition** (DF cas particulier de DMV)

- $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m \Rightarrow A_1, A_2, \dots, A_n \twoheadrightarrow B_1, B_2, \dots, B_m$

GÉNÉRALISATION DIRECTE DU PATRON DE DÉCOMPOSITION



EXEMPLE : FILM



EXEMPLE (SUITE)

Table <i>Film</i>		
titre	nomProducteur	nomActeur
La vie est belle	Elda Ferri	Roberto Benigni
La vie est belle	Elda Ferri	Nicoletta Braschi
La vie est belle	Elda Ferri	Giorgio Cantarini
La vie est belle	Gianluigi Braschi	Roberto Benigni
La vie est belle	Gianluigi Braschi	Nicoletta Braschi
La vie est belle	Gianluigi Braschi	Giorgio Cantarini
Patch Adams	Barry Kemp	Robin Williams
Patch Adams	Barry Kemp	Monica Potter
Patch Adams	Michael Farrell	Robin Williams
Patch Adams	Michael Farrell	Monica Potter
Patch Adams	Marvin Minoff	Robin Williams
Patch Adams	Marvin Minoff	Monica Potter

Table <i>ActeurFilm</i>	
titre	nomActeur
La vie est belle	Roberto Benigni
La vie est belle	Giorgio Cantarini
La vie est belle	Nicoletta Braschi
Patch Adams	Robin Williams
Patch Adams	Monica Potter

Table <i>ProducteurFilm</i>	
titre	nomProducteur
La vie est belle	Elda Ferri
La vie est belle	Gianluigi Braschi
Patch Adams	Barry Kemp
Patch Adams	Michael Farrell
Patch Adams	Marvin Minoff

CINQUIÈME FORME NORMALE (FN5)

Source: Marc Frappier (UdeS), IFT187



DÉFINITION DE 5FN

- Une relation E est en cinquième forme normale ssi E ne peut être obtenue par une jointure de relations E_1, \dots, E_n telle que l'une des E_i n'est pas une clé de E
- la quatrième forme normale est un cas particulier de 5FN; nous omettons sa définition

EXEMPLE DE 5FN

- les relations suivantes sont en 5FN

- | | |
|--------------------------|------------------------|
| – <i>cours</i> | – <i>professeur</i> |
| – <i>prealablesCours</i> | – <i>competence</i> |
| – <i>accessibilite</i> | – <i>disponibilite</i> |
| – <i>groupeCours</i> | – <i>etudiant</i> |
| – <i>inscription</i> | |

CONTRE-EXEMPLE DE 5FN

- offreDeCours(sigle, session, matricule) représente le fait qu'un professeur peut enseigner le cours à une session donnée
- cette relation peut être obtenue par la jointure des 3 relations suivantes:
 - disponibilité(matricule, session)
 - compétence(matricule, sigle)
 - accessibilite(sigle, session)

CONTRE-EXEMPLE DE 5FN

<i>accessibilité</i>	
sigle	session
IFT286	H01
IFT286	E01



<i>disponibilité</i>	
matricule	session
1	E01
2	E01

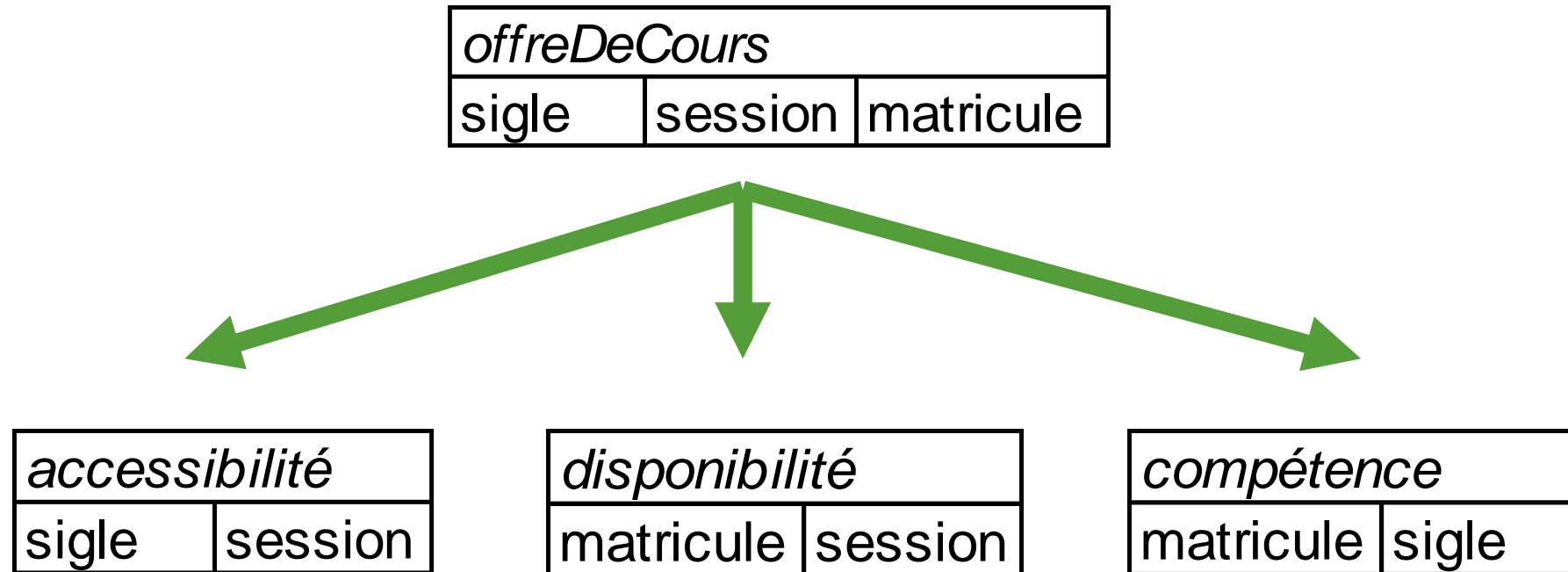


<i>compétence</i>	
matricule	sigle
1	IFT286
2	IFT286

=

<i>offreDeCours</i>		
sigle	session	matricule
IFT286	E01	1
IFT286	E01	2

NORMALISATION EN 5FN



FORMES NORMALES — RÉSUMÉ

- **FN1:** Pas d'attributs multivalués
- **FN2:** FN1 + Pas de dépendances partielles à la clé
- **FN3:** FN2 + Pas de dépendances vers une colonne non-clé
- **FNBC:** FN3 + Une colonne non-clé n'introduit pas de dépendances vers une partie de la clé
- **FN4:** FNBC + Pas de dépendances multivaluées
- **FN5:** FN4 + Toute dépendance de jointure dans la table découle des clés candidates de la table

RECETTE POUR UNE NORMALISATION FACILE

- Il est généralement assez aisé de créer des tables normalisée en FN4 en respectant certaines règles simples :

- Créer des tables à thème unique.

Ex.: une table Clients ne devrait pas avoir de colonnes autres que des colonnes qui décrivent directement un client (i.e. pas les factures, etc.)

- Éviter des clés primaires composées et favoriser des clés artificielles numériques
 - Éviter toute forme de dépendances multivaluées : créer une nouvelle table qui contiendra uniquement chacun des champs multivalués

LES LIMITES DE LA NORMALISATION

- La normalisation permet d'éviter certaines anomalies et des redondances
- La normalisation introduit une complexité au niveau du modèle relationnel à travers les décompositions successives et des associations introduites entre les tables.
- La normalisation n'est pas une fin en soi: On ne vise pas toujours absolument FN6 – On peut accepter (et tenir compte!) de certaines anomalies au profit de la performance
→ Dénormalisation...
- Cependant, dans la majorité des cas, les tables devraient être au moins en FNBC.

NORMALISATION



FN1

FN2

FN3

